

Research Article

Traffic Flow Prediction during the Holidays Based on DFT and SVR

Xianglong Luo^{1,2}, Danyang Li² and Shengrui Zhang¹

¹*School of Highway, Chang'an University, Xi'an 710064, China*

²*School of Information Engineering, Chang'an University, Xi'an 710064, China*

Correspondence should be addressed to Xianglong Luo; xlluo@chd.edu.cn

Received 15 May 2018; Revised 6 September 2018; Accepted 2 October 2018; Published 20 January 2019

Academic Editor: Grigore Stamatescu

Copyright © 2019 Xianglong Luo et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

With the implementation of the freeway free policy during the holidays, traffic congestion in the freeway becomes a common phenomenon. In order to alleviate traffic pressure, traffic flow prediction during the holidays has become a problem of great concern. This paper proposes a hybrid prediction methodology combining discrete Fourier transform (DFT) with support vector regression (SVR). The common trend in the traffic flow data is extracted using DFT by setting an appropriate threshold, which is predicted by extreme extrapolation of the historical trend. The SVR method is applied to predict the residual series. The experimental results with measured data collected from the toll stations in Jiangsu province of China show that the proposed algorithm has higher accuracy compared with the traditional method, and it is an efficient method for traffic flow prediction during the holidays.

1. Introduction

With the rapid development of the social economy, the present massive road infrastructure still fails to meet people's traveling demands [1–5] and traffic congestion has become a common phenomenon in the freeway. Especially, freeways often turn into parking lots during the holidays. To solve the problem of traffic congestion during the holidays, intelligent transportation systems (ITS) have been widely implemented. Traffic flow prediction is fundamental for ITS, and it requires the following preconditions: traffic guidance, planning, and control. Traffic flow is affected by people's willingness to go to a trip, the weather, and many other factors. The implementation of some policies also results in more dramatic changes in traffic flow, such as free charge for cars under 7 seats during important holidays in China, which makes traffic flow prediction more difficult. Moreover, people's travel is more for leisure and tourism during the holidays, which makes some changes in the composition of traffic flow. The passenger car flow increases obviously, while the truck flow decreases to some extent. Especially, traffic flow around

scenic spots is increasing dramatically, and occasional traffic congestions often occur, which results in traffic flow data becoming more stochastic. Traffic flow distribution during the holidays is obviously different from those during workdays. However, most of the existing prediction models are mainly aimed at workdays, and there are relatively few prediction models aimed at the holidays. To improve transportation operation efficiency, accurate traffic flow prediction during the holidays has become a problem of great concern.

Traffic flow prediction is a complex issue. A large number of algorithms have been proposed in the past few years. Existing prediction methods can be basically classified into the following three categories: parametric approach, nonparametric approach, and hybrid approach. The parametric approach includes the historical average (HA) method, autoregressive moving average method (ARIMA) [6, 7], seasonal autoregressive integrated moving average method (SARIMA) [8–10], and Kalman filter (KF) [11, 12]. The nonparametric approach includes artificial neural networks (ANNs) [13–17], *k*-nearest neighbor (KNN) [18–22], support vector regression (SVR) [23, 24], and the

Bayesian model [25, 26]. The hybrid approach mainly combines the parametric approach with the nonparametric approach [27–35]. Vlahogianni et al. [36] summarized existing traffic flow prediction algorithms, and readers interested in details of the models that apply to the field of traffic prediction could refer to their review paper as a reference.

Traffic flow data presents a certain regularity, which makes it possible to predict the traffic flow. In order to improve the accuracy of traffic flow prediction, it is a very effective method to decompose the traffic flow data into different components. There are various methods to retrieve the different components in traffic flow data, including the simple average trend [37], principle component analysis [38], empirical mode decomposition (EMD) [39, 40], wavelet methods [41], spectral analysis method [42, 43], and the low-pass filter [44]. Since traffic flow data is characterized by stochastic and highly nonlinear patterns during the holidays, the existing traffic flow decomposition methods cannot achieve good performance, and it is difficult to apply traditional prediction methods to traffic prediction. Due to the impact of the economic environment, the traffic flow data is almost increasing year by year. Although the annual growth rate is not the same and there is a certain stochastic fluctuation, the trend component of traffic flow data changes under the same law but slightly increases year by year. In order to improve the accuracy of the traffic flow prediction, it is very necessary to extract the trend component in traffic flow data. Li et al. [45, 46] have discussed the trend modeling for traffic time series and explained the benefit of detrending.

Although Dai et al. [47] have specifically studied the traffic flow during the holidays, they focused on the investigation of holiday traffic flow from the spatial and temporal aspects. The ANN model was used to predict the traffic flow, but this method did not aim at the characteristics of traffic flow during the holiday, it still belonged to the traditional prediction model, and it could not meet the requirement of practical engineering. This paper proposes a hybrid prediction methodology combining discrete Fourier transform (DFT) with support vector regression (SVR). Traffic flow time series are decomposed into a common trend and a high-frequency residual time series. The common trend is extracted with the DFT algorithm through setting the appropriate threshold. The common trend reflects the endogenous characteristics, which are approximately stationary. By extreme extrapolation of the historical trend, the common trend of traffic flow time series is predicted. The residual series, which includes fluctuations and bursts, reflects environment-dependent characteristics. As shown in Chen et al. [48] and Li et al. [49], the existence of bursts brings difficulties for prediction models. This paper proposes a method to identify and preprocess bursts, and the prediction of the residual series is estimated by SVR. Finally, the traffic flow in the next holidays is predicted by combining the prediction result of the common trend with the residual series. The tested results with traffic data collected from the toll stations in Jiangsu province of China, show that proposed algorithm has superior accuracy.

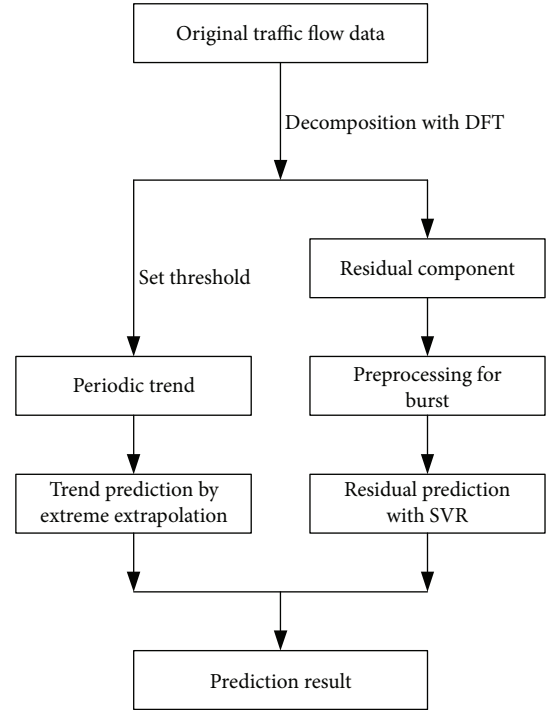


FIGURE 1: Prediction flowchart for the proposed method.

The rest of this paper is organized as follows. Section 2 gives details on a hybrid traffic prediction method based on hybrid DFT and SVR. In Section 3, the dataset used for our numerical experiments is introduced, and the results with different prediction algorithms are presented. Finally, conclusions and the direction for future research are given in Section 4.

2. Methodology

Traffic flow data changes more violently and stochastically, which makes predictions more difficult. In this paper, a hybrid prediction method during the holidays is proposed based on DFT and SVR. The traffic flow data is composed of the common trend and high-frequency residual time series. The traffic flow time series are transformed from the time domain into the frequency domain with DFT, and the trend is extracted through setting the appropriate threshold. Except for the natural growth of the traffic flow during the holiday, the change of the trend component is basically the same. Therefore, the trend component is predicted by extreme extrapolation of the historical trend. For the residual, the fluctuation and burst are defined at first. The mean and variance of the fluctuation are stable, but the burst has great randomness. The burst in the residual component is preprocessed, and then the residual is predicted with SVR. The final prediction result can be obtained by combining the trend with residual prediction result. Our approach is summarized in Figure 1.

2.1. Decomposition of Traffic Flow Time Series with DFT. The purpose of traffic flow time series decomposition is to extract the trend and residual, which can improve the prediction

accuracy. Suppose the sampled traffic flow data on D day holidays can be written as a time series:

$$X_N = \begin{bmatrix} x_{t_1}, x_{t_1+\Delta t}, x_{t_1+2\Delta t}, \dots, x_{t_1+(n-1)\Delta t}, \dots, \\ x_{t_D}, x_{t_D+\Delta t}, x_{t_D+2\Delta t}, \dots, x_{t_D+(n-1)\Delta t} \end{bmatrix}, \quad (1)$$

where Δt is the sample time interval, n is the data size in a day, and if Δt is 1h, then $n = 24$. $N = nD$ is the sample data length on D day holidays.

The spectrum of the time series X_N of the traffic data by DFT can be represented as

$$\hat{x}_k = \sum_{m=0}^{N-1} x_m e^{-j2\pi km/N}, \quad 0 \leq k \leq N-1, \quad (2)$$

where x_m is the m th sample of X_N , \hat{x}_k is the frequency spectrum of x_m , and j is the imaginary unit.

Set a threshold α to remove the high-frequency variations, and the common trend of traffic flow data X_T can be extracted.

$$\tilde{x}_k = \begin{cases} \hat{x}_k, & |\hat{x}_k| > \alpha, \\ 0, & |\hat{x}_k| \leq \alpha, \end{cases} \quad (3)$$

$$x_{T_m} = \frac{1}{N} \sum_{k=0}^{N-1} \tilde{x}_k e^{j2\pi km/N}, \quad 0 \leq m \leq N-1, \quad (4)$$

where x_{T_m} is the m th sample value of the common trend X_T .

Then, traffic flow data is decomposed into the sum of the common trend X_T and the residual X_r :

$$X_N = X_T + X_r. \quad (5)$$

2.2. Prediction of the Common Trend. The trend component represents the regular variations of traffic flow data. Commuters are more and more conscientious of the ecological impacts of their trip, and a change of trend in forthcoming years has attracted more attention. For holidays, the trend component of traffic flow data changes under the same law but slightly increases year by year. The increment is also basically the same, although there is a little difference in western countries. Therefore, the trend component is predicted by extreme extrapolation of the historical trend in this paper.

Suppose that $\bar{X}_{TM} = [X_{T1}, X_{T2}, \dots, X_{TM}]$ is a common trend in consecutive M years, the maximum value $\bar{X}_{T_{\max}}$ of the trend component in M year can then be represented as

$$\begin{aligned} \bar{X}_{T_{\max}} &= [\max(X_{T1}), \max(X_{T2}), \dots, \max(X_{TM})] \\ &= [x_{T1_{\max}}, x_{T2_{\max}}, \dots, x_{TM_{\max}}]. \end{aligned} \quad (6)$$

The increment of the maximum value $\Delta \bar{x}_{T_{\max}}$ can be computed with (7) as

$$\Delta \bar{X}_{T_{\max}} = \left[\frac{x_{T2_{\max}}}{x_{T1_{\max}}}, \frac{x_{T3_{\max}}}{x_{T2_{\max}}}, \dots, \frac{x_{TM_{\max}}}{x_{T(M-1)_{\max}}} \right]. \quad (7)$$

The regression analysis is carried out on $\Delta \bar{X}_{T_{\max}}$, and the regression equation is obtained as

$$C_{\Delta x_T} = \beta_1 t + \beta_2, \quad (8)$$

where β_1 and β_2 are regression coefficients, t is the time variable, and $C_{\Delta x_T}$ is the ratio coefficient of the trend component maximum value compared to that of last year.

The predicted trend component in $M+1$ year is calculated directly with observed traffic flows in consecutive M years.

$$X_{T(M+1)} = C_{\Delta x_T} X_{TM}, \quad (9)$$

where $C_{\Delta x_T}$ is the ratio coefficient calculated by (8) and x_{TM} is the trend component in M year.

2.3. Prediction of the Residual Component

2.3.1. Prediction Preprocessing. The residual series includes fluctuations and bursts. Suppose that the standard deviation σ is obtained from the square root of the variance for the entire residual time series. As demonstrated in Li et al. [45], a point is defined as a “burst” if the entire residual component is twice larger than the standard deviation σ . Most existing prediction models assume that model inputs have a certain smooth mapping relationship with model outputs. Therefore, the predicted value is near the positive and negative variance σ away from the trend component. That is, existing prediction models have a difficulty in forecasting the bursts. The intraday trend is an M -shaped curve that represents the regular variations of daily traffic flow. The morning and evening rush hours result in the two peaks, while one shallow valley and one deep valley, respectively, appear at noon and midnight. This paper extracts the trend component using DFT, which results in the M -shaped curve being smoothed into a number of sine functions, and traffic flow value being slightly enlarged at noon. In order to improve the accuracy of prediction, a strategy is introduced to preprocess the residual component.

$$x_r = \begin{cases} 2\sigma, & x_r > 2\sigma, \\ x_r, & x_r < -2\sigma. \end{cases} \quad (10)$$

In the expression in (10), the positive bursts are preprocessed only twice the standard variance σ , and the negative bursts are not changed, which is different from the existing processing methods. The experiments show that the processing algorithm can help suppress the influence of burst disturbance and improve prediction performance.

2.3.2. Support Vector Regression (SVR). SVR is widely used to deal with nonlinear problems and large-scale prediction

algorithms. In this paper, the SVR method is applied to predict the preprocessed residual series. For a given training set $T = \{(x_i, y_i), i = 1, 2, \dots, l, x_i \in R^n, y_i \in R\}$, where x_i is a dimensional input variable, y_i is the corresponding output value, and l represents the size of the training data.

Through a nonlinear mapping function $\phi(\cdot)$, x_i is mapped to a high-dimensional feature space, in which the optimal decision function is constructed as follows:

$$f(x) = w^T \phi(x) + b, \quad w \in R^n, b \in R, \quad (11)$$

where w is the weight vector and b is the bias value. SVR minimizes the following objective function:

$$\begin{aligned} \min \quad & \left[\frac{1}{2} \|w\|^2 + C \sum_{i=1}^l \xi_i + \xi_i^* \right] \\ \text{s.t.} \quad & \begin{cases} (wx_i + b) - y_i \leq \varepsilon + \xi_i \\ y_i - (wx_i + b) \leq \varepsilon + \xi_i^* \\ \xi_i \geq 0, \xi_i^* \geq 0, \varepsilon \geq 0, \end{cases} \end{aligned} \quad (12)$$

where $1/2\|w\|^2$ is the expression for the complexity of the decision function, C is a penalty coefficient, and ξ_i is the

upper training error (ξ_i^* is lower) subject to the ε insensitive factor.

This optimization problem can be transformed into a dual problem, and its solution is given by (13), where a_i^* , a_i are the Lagrange multipliers that can be acquired by solving the dual problem and $K(x_i, x)$ is the Mercer kernel function that equals the inner product of $\phi(x_i)$ with $\phi(x)$.

$$\begin{aligned} f(x) &= \sum_{i=1}^l (a_i^* - a_i) K(x_i, x) + b, \\ \text{s.t.} \quad & 0 \leq a_i^* \leq C, 0 \leq a_i \leq C. \end{aligned} \quad (13)$$

The most frequently used Mercer kernel functions are instance sigmoid, polynomial, and radial basis function (RBF). RBF is a common choice for the kernel function because of the need to set very few parameters and excellent overall performance. So, RBF is selected as the kernel function in this paper.

2.4. Proposed Prediction Model. Suppose the sampled traffic flow data on D day holidays in consecutive M years can be written as the matrix X_{MN} :

$$X_{MN} = \begin{bmatrix} x_{t_{11}}, & x_{t_{11}+\Delta t}, & x_{t_{11}+2\Delta t} & \cdots & x_{t_{11}+(n-1)\Delta t}, & \cdots, & x_{t_{1D}}, & x_{t_{1D}+\Delta t}, & x_{t_{1D}+2\Delta t} & \cdots & x_{t_{1D}+(n-1)\Delta t} \\ x_{t_{21}}, & x_{t_{21}+\Delta t}, & x_{t_{21}+2\Delta t} & \cdots & x_{t_{21}+(n-1)\Delta t}, & \cdots, & x_{t_{2D}}, & x_{t_{2D}+\Delta t}, & x_{t_{2D}+2\Delta t} & \cdots & x_{t_{2D}+(n-1)\Delta t} \\ \vdots & \vdots & \vdots & & \vdots & & \vdots & \vdots & \vdots & & \vdots \\ x_{t_{M1}}, & x_{t_{M1}+\Delta t}, & x_{t_{M1}+2\Delta t} & \cdots & x_{t_{M1}+(n-1)\Delta t}, & \cdots, & x_{t_{MD}}, & x_{t_{MD}+\Delta t}, & x_{t_{MD}+2\Delta t} & \cdots & x_{t_{MD}+(n-1)\Delta t} \end{bmatrix}, \quad (14)$$

where Δt is the sample time interval, n is the data size in a day, and $N = nD$ is the sample data length on D day holidays.

According to (3), (4), and (5), the traffic flow data in consecutive M years are decomposed into the sum of the trend components and the residuals.

$$X_{MN} = \bar{X}_T + X_R = \begin{bmatrix} X_{T1} \\ X_{T2} \\ \vdots \\ X_{TM} \end{bmatrix} + \begin{bmatrix} X_{R1} \\ X_{R2} \\ \vdots \\ X_{RM} \end{bmatrix}, \quad (15)$$

where \bar{X}_T are the trend components in consecutive M years and X_R are the residuals.

The predicted traffic flow during the holidays in $M + 1$ year is calculated directly by adding the trend with residual prediction results.

$$\hat{X}_{(M+1)} = f_1(\bar{X}_T) + f_2(X_R), \quad (16)$$

where $f_1(\bar{X}_T)$ is the prediction result of the trend component with (9) and $f_2(X_R)$ is the prediction result of the residual component with SVR.

3. Case Study

3.1. Data Description. Jiangsu is a highly developed province in China. In this paper, the data used to evaluate the performance of the proposed model are collected from the toll stations in Jiangsu province on Tomb-sweeping Day and National Day from 2011 to 2015. The length of the traffic flow data interval has an important influence on the prediction results. An excessively short time interval tends to lead to the instability of the prediction model, but an interval that is too long will lose the inherent characteristics of traffic flow data. As shown in Ma et al. [50], the prediction performance improves as the data aggregation level increases due to less data fluctuation as the time interval becomes longer. The prediction accuracy reaches its maximum when the aggregation level reaches one hour. Therefore, we selected one hour as the sampling time, which can also meet engineering management requirements. There are total of 1200 data points (50

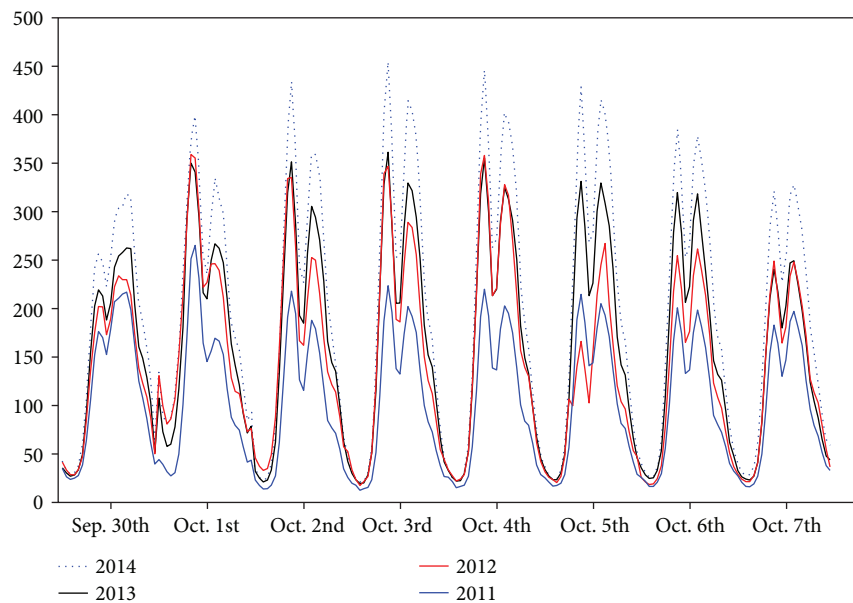


FIGURE 2: The traffic flow on National Day from 2011 to 2014.

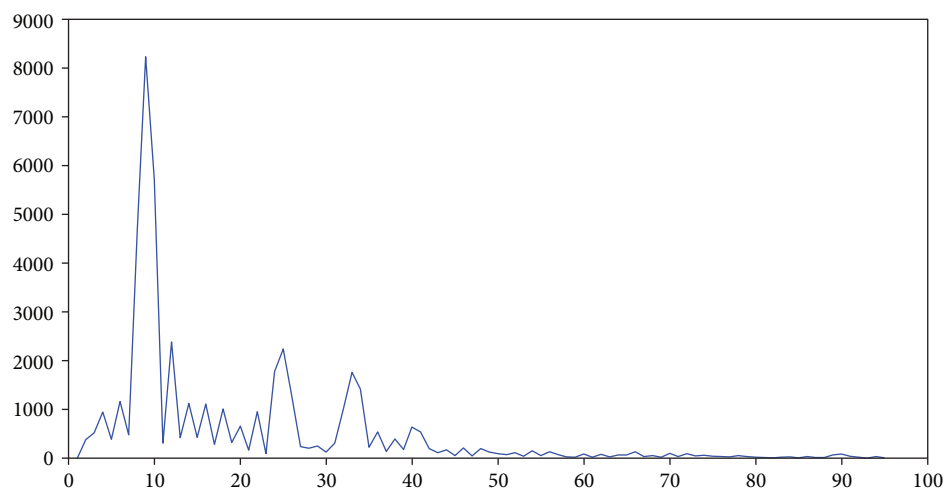


FIGURE 3: The spectrum of traffic flow on National Day in 2011.

days from 2011 to 2015) in the dataset. The data are divided into two data sets. The first 960 data points (40 days from 2011 to 2014) are used as the training sample, while the remaining 240 data points (10 days in 2015) are employed as the testing sample for measuring the prediction performance of the proposed model. The traffic flow on National Day from 2011 to 2014 is shown in Figure 2. From Figure 2, we can see that the trend of change of the traffic flow is basically consistent every year and the traffic flow value increases around 10% year by year. The traffic flow data are similar during the whole holiday in each year except for the first and end day of the holiday, which shows a sharply increased and decreased flow. The hourly distribution of traffic flow in each day has the same trend with a high value in the daytime and a low value at night. It shows remarkable similarity and strong periodic features, but it also reveals the stochastic properties. It is

necessary to decompose the traffic flow data in order to reveal its potential characteristics.

3.2. Decomposition of Traffic Flow Time Series with DFT. The traffic flow data increases year by year in Figure 2, but has similar patterns during the whole holiday from 2011 to 2014. It has been pointed out in literature that the prediction accuracy is greatly affected because of the existence of several trends, and the trends should be removed to improve the prediction performance. DFT is a powerful analytical method to extract the periodic component of the signal. This paper decomposes the traffic flow data into a trend and a residual component with DFT. The spectrum of the traffic flow on National Day in 2011 is shown in Figure 3, and three obvious components can be seen. The sampling period of the testing dataset is one hour, then $n = 24$, the sample data length on D day

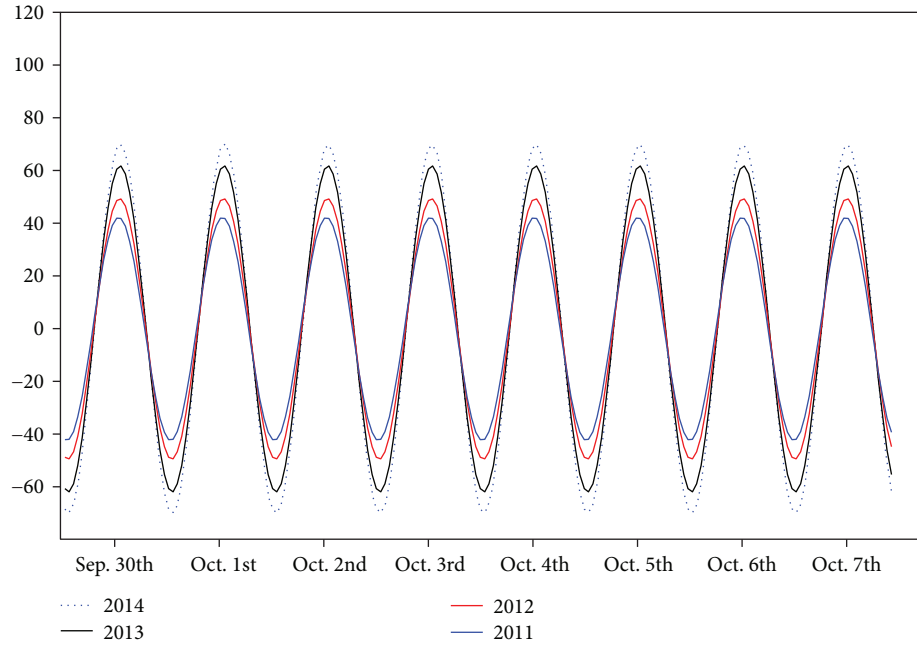


FIGURE 4: The trend components on National Day from 2011 to 2014.

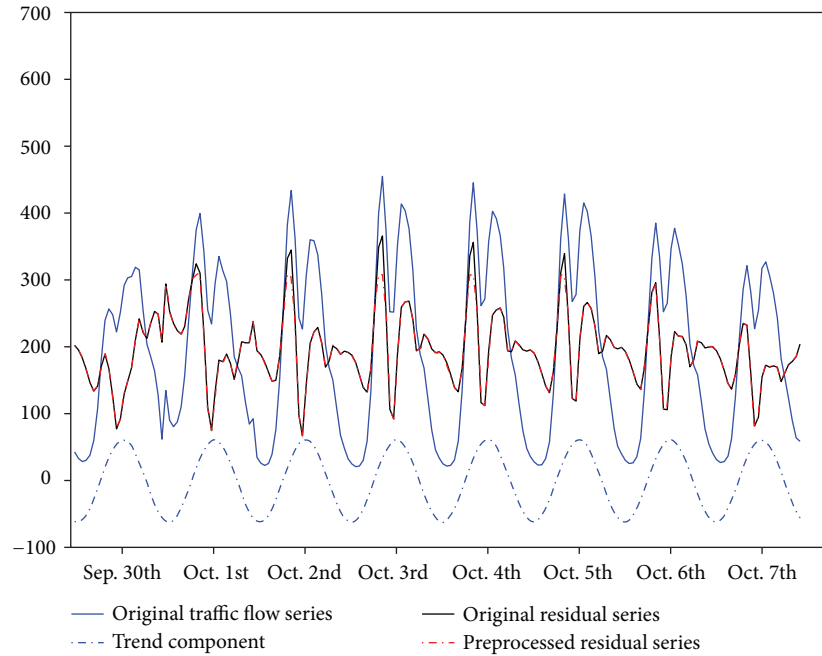
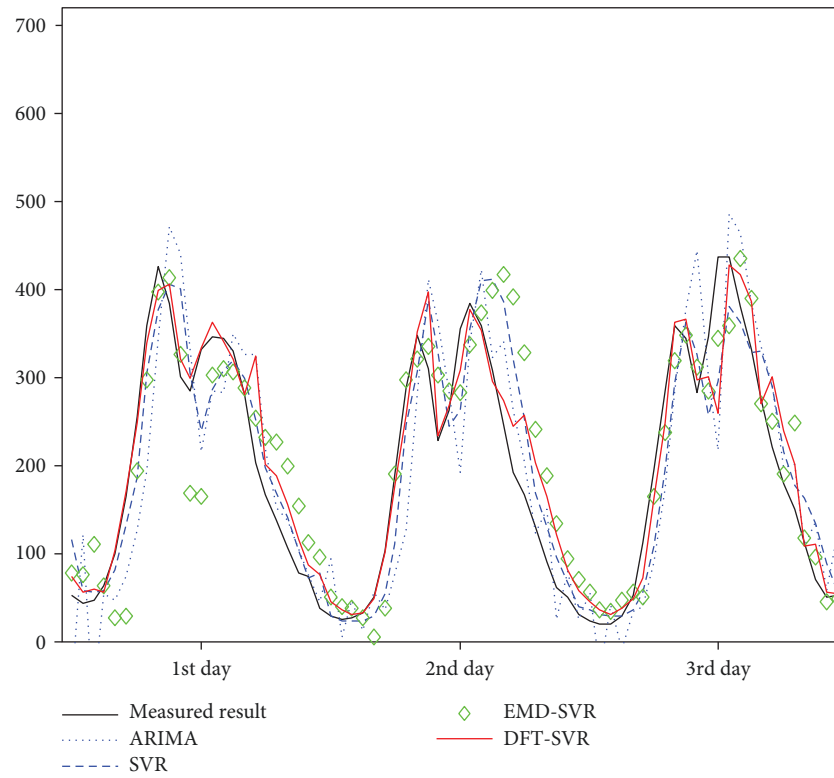


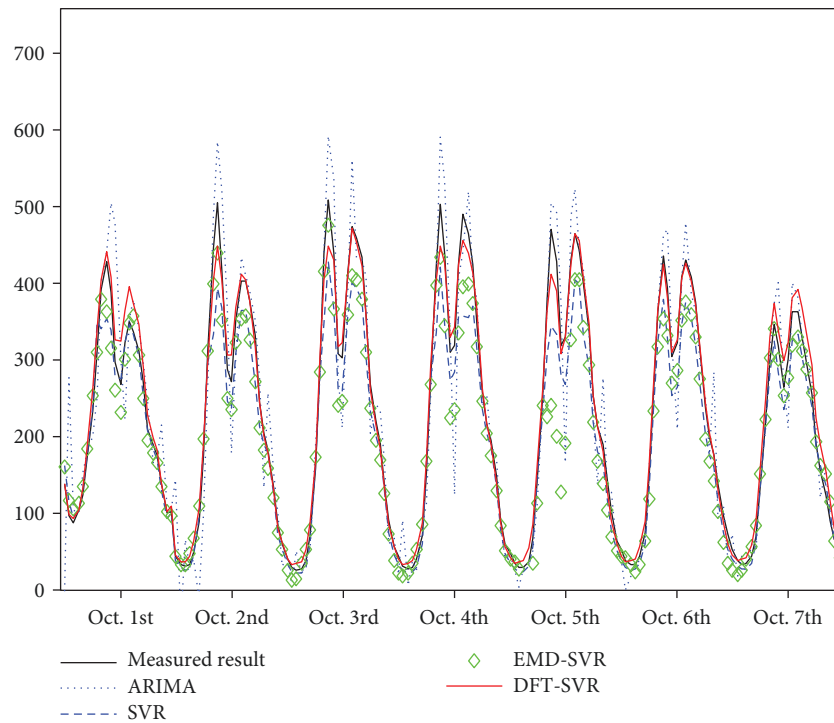
FIGURE 5: The original traffic flow series, trend component, residual series, and preprocessed residual series on National Day in 2014.

holidays is $N = nD$. If the data length is the power of 2, we use DTF directly. Otherwise we will fill the data to make it a power of 2 by adding 0 in the end of the original traffic data, which will not change the frequency resolution and also conform to the calculation condition for DFT. From Figure 3, we can see that there are several peaks in the traffic flow spectrum and the intervals between the peaks are large enough to meet the frequency resolution requirements. The main purpose of DFT is to extract the trend

components of traffic flow sequences. This indicates that we need to retain the low-frequency components and suppress high-frequency components. Through experiments, we can extract the trend components of traffic flow data effectively by keeping only the highest peak value component. However, the highest peak is not really an impulse function, so we must set a large threshold to suppress other frequency components. When the threshold α is set 0.7 times of the maximum frequency spectrum value,



(a)



(b)

FIGURE 6: The prediction results by DFT-SVR, ARIMA, SVR, and EMD-SVR (a) on Tomb-sweeping Day and (b) on National Day.

the trend components of the traffic flow sequences can be extracted efficiently according to (3) and (4). The trend components on National Day from 2011 to 2015 are shown in Figure 4. Compared to Figure 2, the law of trend

components is more obvious year by year, so it becomes easier to predict the trend component in traffic flow. The traffic flow sequence minus the trend component and the residual series are obtained.

3.3. Preprocessing of the Residual Series. The residual series includes fluctuations and bursts, and it represents the stochastic component in the traffic flow, which is affected by the weather, traffic incidence, roadway conditions, and other factors. Especially, because of the presence of large bursts, the accuracy of traffic flow prediction is greatly affected. In order to improve the accuracy of the prediction, the residual series are preprocessed according to (10). The original traffic flow series, trend component, residual series, and preprocessed residual series on National Day in 2014 are shown in Figure 5.

3.4. Prediction Results and Performance Comparison

3.4.1. Measuring Performance of the Proposed Prediction Method. Several standard evaluation measurements are adopted in the experiments to evaluate the proposed method, including the mean absolute error (MAE), mean absolute percentage error (MAPE), and root mean square error (RMSE). Although these indicators are used to evaluate the prediction performance, each indicator has its emphasis on evaluation. In order to evaluate the performance of the proposed model accurately, we have adopted three indicators at the same time to represent the difference between the actual value and the predicted value.

$$\begin{aligned} \text{MAE} &= \frac{1}{N} \sum_{i=1}^N |\hat{x}_i - x_i|, \\ \text{MAPE} &= \frac{1}{N} \sum_{i=1}^N \left| \frac{\hat{x}_i - x_i}{x_i} \right| \times 100\%, \\ \text{RMSE} &= \sqrt{\frac{1}{N} \sum_{i=1}^N (\hat{x}_i - x_i)^2}, \end{aligned} \quad (17)$$

where \hat{x}_i is the prediction value, x_i is the actual observed value, and N is the length of the test series.

3.4.2. Prediction Results. In this section, the traffic flows on Tomb-sweeping Day and National Day in 2015 are predicted with the proposed DFT and SVR model using the traffic flow data from 2011 to 2014. Figures 6(a) and 6(b), respectively, exhibit the predicted and real traffic flow on Tomb-sweeping Day and National Day by 4 different models: DFT-SVR, ARIMA, SVR, and EMD-SVR. From this figure, it is apparent that the proposed DFT and SVR model has the best prediction performance, and the prediction value of the proposed method is identical to the real value during most of the hours, especially during the rush hours. The prediction results by ARIMA and EMD-SVR are worse than that of the proposed model. Because holidays are longer on National Day compared to Tomb-sweeping Day, the traffic flow is relatively smooth on National Day. However, Tomb-sweeping Day is a traditional holiday for tomb-sweeping, and traffic flow data during this holiday is very high with a more stochastic characteristic. The prediction accuracy is higher on National Day than on Tomb-sweeping Day in Figure 6.

TABLE 1: The prediction performance on Tomb-sweeping Day.

	MAE	MAPE (%)	RMSE
ARIMA	63.8955	53.2863	0.7967
SVR	43.8453	29.3454	0.3801
EMD-SVR	50.9304	43.6135	0.5995
DFT-SVR	28.004	12.1147	0.3700

TABLE 2: The prediction performance on National Day.

	MAE	MAPE (%)	RMSE
ARIMA	37.0084	26.7927	0.4788
SVR	29.0494	13.3154	0.1496
EMD-SVR	32.9437	28.4532	0.4791
DFT-SVR	14.0734	9.6237	0.1386

The prediction performances are shown on Tomb-sweeping Day and National Day in Tables 1 and 2. It can be seen that the proposed DFT-SVR method has better prediction performance than ARIMA, SVR, and EMD-SVR on both Tomb-sweeping Day and National Day. All algorithms have better prediction accuracy on long holidays than on short holidays, and the average prediction accuracy is higher at 15.04% on National Day than on Tomb-sweeping Day. The prediction performance of the proposed DFT-SVR method improved at least by 3.69% on National Day, but it improved by at least 17.23% on Tomb-sweeping Day, indicating that the proposed method has more advantages for dealing with stochastic traffic flow data.

4. Conclusions

This paper presents a traffic flow prediction method for the holidays based on DFT and SVR. The trend components are extracted from holiday traffic flow data with DFT, and the trend can be predicted by extreme extrapolation with historical data. The residual component is further analyzed by a preprocessing method and SVR. The final prediction result is obtained by combining the trend with the residual prediction result. Real traffic data, collected from toll stations in Jiangsu province from 2011 to 2015, are used to evaluate prediction performances of the proposed DFT-SVR model. The test results show that the proposed DFT-SVR model has the best performance compared with ARIMA, SVR, and EMD-SVR, and prediction accuracy is higher on long holidays than on short holidays. Since the traffic flow data during the holidays is affected by weather and other factors, the impact of weather factors on traffic flow will also be further studied, so as to improve the prediction accuracy.

Data Availability

The data used in this paper were collected from the toll stations in Jiangsu province on Tomb-sweeping Day and National Day from 2011 to 2015. The data are supplied by the JiangSu Expressway Network Operation & Management Center in China. Because the data involve license and

commercial confidentiality, these cannot be used freely. If any researcher requests for these data, he can send an email to xlluo@chd.edu.cn.

Conflicts of Interest

The authors declare that there is no conflict of interest regarding the publication of this paper.

Acknowledgments

This research was partly supported by the National Key R&D Program of China (2018YFC0808706) and the National Natural Science Foundation of China (Grant no. 5157081053). The authors are also grateful to the JiangSu Expressway Network Operation & Management Center for providing the data.

References

- [1] Y. Li, S. Xu, H. Liu, E. Ma, and L. Wang, "Displacement and stress characteristics of tunnel foundation in collapsible loess ground reinforced by jet grouting columns," *Advances in Civil Engineering*, vol. 2018, Article ID 2352174, 16 pages, 2018.
- [2] J. Lai, S. He, J. Qiu et al., "Characteristics of seismic disasters and aseismic measures of tunnels in Wenchuan earthquake," *Environmental Earth Sciences*, vol. 76, no. 94, pp. 1–19, 2017.
- [3] J. Qiu, T. Yang, X. Wang, L. Wang, and G. Zhang, "Review of the flame retardancy on highway tunnel asphalt pavement," *Construction and Building Materials*, vol. 195, pp. 468–482, 2019.
- [4] J. Lai, K. Wang, J. Qiu, F. Niu, J. Wang, and J. Chen, "Vibration response characteristics of the cross tunnel structure," *Shock and Vibration*, vol. 2016, Article ID 9524206, 16 pages, 2016.
- [5] J. Lai, X. Wang, J. Qiu, J. Chen, Z. Hu, and H. Wang, "Extreme deformation characteristics and countermeasures for a tunnel in difficult grounds in southern Shaanxi, China," *Environmental Earth Sciences*, vol. 77, no. 706, Article ID 9183296, pp. 1–19, 2018.
- [6] C. Han, S. Song, and C. H. Wang, "A real-time short-term traffic flow adaptive forecasting method based on ARIMA model," *Acta Simulata Systematica Sinica*, vol. 16, no. 7, pp. 1530–1535, 2004.
- [7] M. C. Tan, S. C. Wong, J. M. Xu, Z. R. Guan, and P. Zhang, "An aggregation approach to short-term traffic flow prediction," *IEEE Transactions on Intelligent Transportation Systems*, vol. 10, no. 1, pp. 60–69, 2009.
- [8] X. Luo, L. Niu, and S. Zhang, "An algorithm for traffic flow prediction based on improved SARIMA and GA," *KSCE Journal of Civil Engineering*, vol. 22, no. 10, pp. 4107–4115, 2018.
- [9] B. M. Williams and L. A. Hoel, "Modeling and forecasting vehicular traffic flow as a seasonal ARIMA process: theoretical basis and empirical results," *Journal of Transportation Engineering*, vol. 129, no. 6, pp. 664–672, 2003.
- [10] G. Shi, J. Guo, W. Huang, and B. M. Williams, "Modeling seasonal heteroscedasticity in vehicular traffic condition series using a seasonal adjustment approach," *Journal of Transportation Engineering*, vol. 140, no. 5, article 04014012, 2014.
- [11] L. L. Ojeda, A. Y. Kibangou, and C. C. de Wit, "Adaptive Kalman filtering for multi-step ahead traffic flow prediction," in *2013 American Control Conference*, pp. 4724–4729, Washington, DC, USA, 2013, IEEE.
- [12] S. V. Kumar, "Traffic flow prediction using Kalman filtering technique," *Procedia Engineering*, vol. 187, pp. 582–587, 2017.
- [13] B. L. Smith and M. J. Demetsky, "Short-term traffic flow prediction: neural network approach," *Transportation Research Record*, vol. 1453, pp. 98–104, 1994.
- [14] T. Zhou, G. Han, X. Xu et al., " δ -Agree AdaBoost stacked auto-encoder for short-term traffic flow forecasting," *Neurocomputing*, vol. 247, pp. 31–38, 2017.
- [15] H. F. Yang, T. S. Dillon, and Y. P. P. Chen, "Optimized structure of the traffic flow forecasting model with a deep learning approach," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 28, no. 10, pp. 2371–2381, 2017.
- [16] T. Ma, Z. Zhou, and B. Abdulhai, "Nonlinear multivariate time-space threshold vector error correction model for short term traffic state prediction," *Transportation Research Part B: Methodological*, vol. 76, pp. 27–47, 2015.
- [17] Z. Ma, G. Luo, and D. Huang, "Short term traffic flow prediction based on on-line sequential extreme learning machine," in *2016 Eighth International Conference on Advanced Computational Intelligence (ICACI)*, pp. 143–149, Chiang Mai, Thailand, 2016, IEEE.
- [18] M. Bernaś, B. Płaczek, P. Porwik, and T. Pamuła, "Segmentation of vehicle detector data for improved k -nearest neighbours-based traffic flow prediction," *IET Intelligent Transport Systems*, vol. 9, no. 3, pp. 264–274, 2015.
- [19] Y. Cai, H. Huang, H. Cai, and Y. Qi, "A K -nearest neighbor locally search regression algorithm for short-term traffic flow forecasting," in *2017 9th International Conference on Modelling, Identification and Control (ICMIC)*, pp. 624–629, Kunming, China, 2017.
- [20] S. Wu, Z. Yang, X. Zhu, and B. Yu, "Improved k -nn for short-term traffic forecasting using temporal and spatial information," *Journal of Transportation Engineering*, vol. 140, no. 7, article 04014026, 2014.
- [21] P. Dell'Acqua, F. Bellotti, R. Berta, and A. De Gloria, "Time-aware multivariate nearest neighbor regression methods for traffic flow prediction," *IEEE Transactions on Intelligent Transportation Systems*, vol. 16, no. 6, pp. 3393–3402, 2015.
- [22] B. Sun, W. Cheng, P. Goswami, and G. Bai, "Short-term traffic forecasting using self-adjusting k -nearest neighbours," *IET Intelligent Transport Systems*, vol. 12, no. 1, pp. 41–48, 2018.
- [23] M. Castro-Neto, Y. S. Jeong, M. K. Jeong, and L. D. Han, "Online-SVR for short-term traffic flow prediction under typical and atypical traffic conditions," *Expert Systems with Applications*, vol. 36, no. 3, pp. 6164–6173, 2009.
- [24] Y. Sun, B. Leng, and W. Guan, "A novel wavelet-SVM short-time passenger flow prediction in Beijing subway system," *Neurocomputing*, vol. 166, pp. 109–121, 2015.
- [25] W. Zheng, D.-H. Lee, and Q. Shi, "Short-term freeway traffic flow prediction: Bayesian combined neural network approach," *Journal of Transportation Engineering*, vol. 132, no. 2, pp. 114–121, 2006.
- [26] Q. P. Wang and H. Sun, "Traffic structure optimization in historic districts based on green transportation and sustainable development concept," *Advances in Civil Engineering*, vol. 2019, Article ID 9196263, 16 pages, 2019.
- [27] Z. Shen, W. Wang, Q. Shen, and Z. Li, "Hybrid CSA optimization with seasonal RVR in traffic flow forecasting,"

- KSII Transactions on Internet and Information Systems*, vol. 11, no. 10, pp. 4867–4887, 2017.
- [28] Z. Hu, K. Du, J. Lai, and Y. Xie, “Statistical analysis of influence of cover depth on loess tunnel deformation in NW China,” *Advances in Civil Engineering*, vol. 2019, Article ID 2706976, 12 pages, 2019.
 - [29] D. Xia, B. Wang, H. Li, Y. Li, and Z. Zhang, “A distributed spatial–temporal weighted model on MapReduce for short-term traffic flow forecasting,” *Neurocomputing*, vol. 179, pp. 246–263, 2016.
 - [30] F. Moretti, S. Pizzuti, S. Panziera, and M. Annunziato, “Urban traffic flow forecasting through statistical and neural network bagging ensemble hybrid modeling,” *Neurocomputing*, vol. 167, pp. 3–7, 2015.
 - [31] J. Lai, J. Qiu, H. Fan et al., “Fiber bragg grating sensors-based in situ monitoring and safety assessment of loess tunnel,” *Journal of Sensors*, vol. 2016, Article ID 8658290, 10 pages, 2016.
 - [32] M. W. Li, D. F. Han, and W. L. Wang, “Vessel traffic flow forecasting by RSVR with chaotic cloud simulated annealing genetic algorithm and KPCA,” *Neurocomputing*, vol. 157, pp. 243–255, 2015.
 - [33] S. D. Oh, Y. J. Kim, and J. S. Hong, “Urban traffic flow prediction system using a multifactor pattern recognition model,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 16, no. 5, pp. 2744–2755, 2015.
 - [34] S. He, L. Su, H. Fan, and R. Ren, “Methane Explosion Accidents of Tunnels in SW China,” *Geomatics, Natural Hazards and Risk*, 2018.
 - [35] Z. Zhang, H. Zhang, Y. Tan, and H. Yang, “Natural wind utilization in the vertical shaft of a super-long highway tunnel and its energy saving effect,” *Building and Environment*, vol. 145, pp. 140–152, 2018.
 - [36] E. I. Vlahogianni, M. G. Karlaftis, and J. C. Golias, “Short-term traffic forecasting: where we are and where we’re going,” *Transportation Research Part C: Emerging Technologies*, vol. 43, pp. 3–19, 2014.
 - [37] W. Min and L. Wynter, “Real-time road traffic prediction with spatio-temporal correlations,” *Transportation Research Part C: Emerging Technologies*, vol. 19, no. 4, pp. 606–616, 2011.
 - [38] L. Qu, J. Hu, L. Li, and Y. Zhang, “PPCA-based missing data imputation for traffic flow volume: a systematical approach,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 10, no. 3, pp. 512–522, 2009.
 - [39] M. C. Chen and Y. Wei, “Exploring time variants for short-term passenger flow,” *Journal of Transport Geography*, vol. 19, no. 4, pp. 488–498, 2011.
 - [40] H. Wang, L. Liu, Z. (S.) Qian, H. Wei, and S. Dong, “Empirical mode decomposition—autoregressive integrated moving average: hybrid short-term traffic speed prediction model,” *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2460, no. 1, pp. 66–76, 2014.
 - [41] D. Boto-Giralda, F. J. Díaz-Pernas, D. González-Ortega et al., “Wavelet-based denoising for traffic volume time series forecasting with self-organizing neural networks,” *Computer-Aided Civil and Infrastructure Engineering*, vol. 25, no. 7, pp. 530–545, 2010.
 - [42] T. T. Tchakian, B. Basu, and M. O’Mahony, “Real-time traffic flow forecasting using spectral analysis,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 13, no. 2, pp. 519–526, 2012.
 - [43] H. Nicholson and C. D. Swann, “The prediction of traffic flow volumes based on spectral analysis,” *Transportation Research*, vol. 8, no. 6, pp. 533–538, 1974.
 - [44] T. Otoshi, Y. Ohsita, M. Murata, Y. Takahashi, K. Ishibashi, and K. Shiimoto, “Traffic prediction for dynamic traffic engineering,” *Computer Networks*, vol. 85, pp. 36–50, 2015.
 - [45] L. Li, X. Su, Y. Zhang, Y. Lin, and Z. Li, “Trend modeling for traffic time series analysis: an integrated study,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 16, no. 6, pp. 3430–3439, 2015.
 - [46] Y. Li, Z. Li, and L. Li, “Missing traffic data: comparison of imputation methods,” *IET Intelligent Transport Systems*, vol. 8, no. 1, pp. 51–57, 2014.
 - [47] H. Dai, F. Wang, Q. Liu, and C. Gong, “Investigation and prediction of traffic flow in holidays in Zhejiang section of Shanghai freeway,” in *2015 International Conference on Transportation Information and Safety (ICTIS)*, pp. 195–201, Wuhan, China, 2015.
 - [48] C. Chen, Y. Wang, L. Li, J. Hu, and Z. Zhang, “The retrieval of intra-day trend and its influence on traffic prediction,” *Transportation Research Part C: Emerging Technologies*, vol. 22, no. 5, pp. 103–118, 2012.
 - [49] L. Li, X. Su, Y. Wang, Y. Lin, Z. Li, and Y. Li, “Robust causal dependence mining in big data network and its application to traffic flow predictions,” *Transportation Research Part C: Emerging Technologies*, vol. 58, pp. 292–307, 2015.
 - [50] X. Ma, H. Yu, Y. Wang, and Y. Wang, “Large-scale transportation network congestion evolution prediction using deep learning theory,” *PLoS One*, vol. 10, no. 3, article e0119044, 2015.

